Formal Properties Of The Metathesaurus: An Update*

Stephanie S. Lipow[1], Keith E. Campbell, MD[2], Nels E. Olson[1],
Mark S. Tuttle[1], Mark S. Erlbaum, MD[1], Lloyd F. Fuller, PhD[1],
David D. Sherertz[1], Stuart J. Nelson, MD[3], William G. Cole, PhD[1,4]
[1]Lexical Technology, Inc., Alameda, CA; [2]Stanford University, Stanford, CA ;
[3]Medical College of Georgia, Augusta, GA; [4]University of Washington, Seattle, WA

Formality will be a key to the cost effective use, maintenance and enhancement of the UMLS Metathesaurus[1]. In 1994, we reported that the Metathesaurus was becoming increasingly formal, despite its dramatic growth.[2] The 1995 Metathesaurus continues this trend. Our report offers updated figures and additional observations regarding the benefits of this formality.

The formality of the Metathesaurus stems from the way that its concepts are related to one another. Thus our first interest is in the relative stability of those concepts. In 1994, we observed that *atoms*, names in a constituent terminology, are being added to the Metathesaurus faster than *concepts*, unique, named meanings. **Figure 1** shows the number of *atoms* and *concepts* in the reviewed portion of the six versions of the Metathesaurus, and the 1995 Metathesaurus continues the trend toward convergence. If this continues for future versions of the Metathesaurus, it may indicate the emergence of an empirical consensus on the identity of biomedical concepts independent of what they are called.

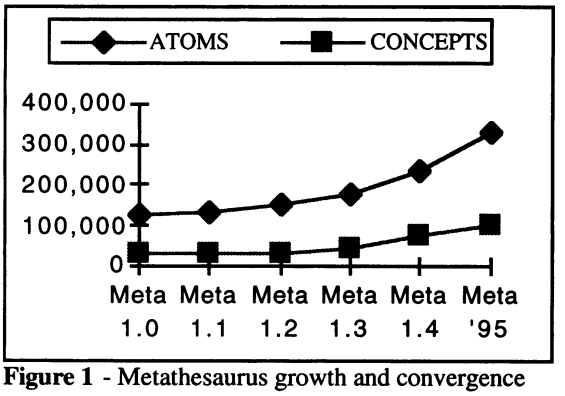

Figure 1 - Metathesaurus growth and convergence

Given a collection of stable concepts, one measure of their formality is their degree of conformance to a traditional "Aristotelian" Hierarchy. This conformance can be defined as the proportion of concepts that have both *genera*, i.e. belong to a hierarchy, and *differentia*, i.e. are distinguishable from their siblings in that hierarchy. *Genera* include "parent" and "broader" relationships as well as the hierarchical relationships resulting from each concept's semantic types. Horizontal ("other") relationships, definitions, associated expressions, and co-occurrences are counted as *differentia*. **Figure 2** shows the total number of *concepts* and the number with both *genera* and *differentia* for the last four versions of the Metathesaurus. Based on these measurements, the 1995 version demonstrates increasingly "Aristotelian" formal properties.

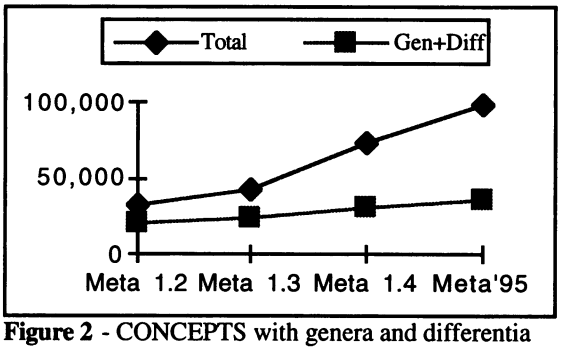

Figure 2 - CONCEPTS with genera and differentia

These formal properties have potentially significant practical benefits. First, they will allow the Metathesaurus to be more productively exploited by computer applications. The Aristotelian connections support both navigation among related terms and aggregation of subordinate terms. For example, Metathesaurus-based applications could "return all patients with an infectious disease," and "return all patients with a hepatic disease," and have patients with "infectious hepatitis" properly included in each set. Second, they will facilitate Metathesaurus maintenance and enhancement. The 1995 Metathesaurus contains concepts from 35 source vocabularies, including nine entirely new sources and a number of updates to existing sources. Relationships and other information about concepts need to be added or adjusted whenever new sources or new versions of existing sources are incorporated. Aristotelian definitions for the to-be-merged concepts can help determine if the relevant terms are synonyms, or not. For example, terms that are not lexically similar such as "liver inflammation" and "hepatitis" may be left unlinked in the Metathesaurus because existing tools would not identify them as potential synonyms. However, if each term had been defined with the genus "inflammation" and the differentia "affecting the liver," then an algorithm can suggest that they might be synonyms.

* Partially supported by National Library of Medicine contract N01-LM-3-3515.

1. Campbell KE. Distributed development of a logic-based controlled medical terminology (Dissertation Proposal). Stanford Univ., 1994.
2. Tuttle MS, et al., Formal properties of the Metathesaurus. SCAMC, 1994:145-149.